

That which is claimed:

1. A method comprising:

identifying a boilerplate element in an article comprising a plurality of elements,
the plurality of elements comprising the boilerplate element and a content element; and
generating an implicit search query comprising a search term, the search term
comprising a term present in the content element.

2. The method of claim 1, wherein identifying the boilerplate element comprises

identifying a common element in a plurality of related articles.

3. The method of claim 1, wherein the common element comprises a copyright notice.

4. The method of claim 1, wherein the common element comprises a term having a

low inverse document frequency measure.

5. The method of claim 1, wherein identifying the boilerplate element comprises

comparing at least one of the plurality of elements to a predetermined list.

6. The method of claim 1, wherein identifying the boilerplate element comprises

analyzing the spatial location at least one of the plurality of elements.

7. The method of claim 1, wherein identifying the boilerplate element comprises

analyzing a navigational element of the article.

8. The method of claim 1, wherein identifying the boilerplate element comprises analyzing a link element of the article.
9. The method of claim 8, wherein analyzing the link element of the article comprises analyzing an address to which the link element refers.
10. The method of claim 1, wherein identifying the boilerplate element comprises analyzing a markup language element proximate to at least one of the plurality of elements.
11. A method comprising:
 - identifying a boilerplate element in an article comprising a plurality of elements,
 - the plurality of elements comprising the boilerplate element and a content element;
 - removing the boilerplate element from the article; and
 - indexing the article.
12. A method comprising:
 - identifying a boilerplate element in an article comprising a plurality of elements,
 - the plurality of elements comprising the boilerplate element and a content element; and
 - determining a weight for each of the plurality of elements based at least in part on whether the element is a boilerplate element.
13. The method of claim 12, further comprising:

receiving a search query;
determining articles relevant to the search query; and
ranking the articles based at least in part on the determined weights.

14. A computer-readable medium on which is encoded program code, the program code comprising:

program code for identifying a boilerplate element in an article comprising a plurality of elements, the plurality of elements comprising the boilerplate element and a content element; and

program code for generating an implicit search query comprising a search term, the search term comprising a term present in the content element.

15. The computer-readable medium of claim 14, wherein program code for identifying the boilerplate element comprises program code for identifying a common element in a plurality of related articles.

16. The computer-readable medium of claim 14, wherein identifying the boilerplate element comprises comparing at least one of the plurality of elements to a predetermined list.

17. The computer-readable medium of claim 14, wherein identifying the boilerplate element comprises analyzing the spatial location at least one of the plurality of elements.

18. The computer-readable medium of claim 14, wherein identifying the boilerplate element comprises analyzing a navigational element of the article.
19. The computer-readable medium of claim 14, wherein identifying the boilerplate element comprises analyzing a link element of the article.
20. The computer-readable medium of claim 16, wherein analyzing the link element of the article comprises analyzing an address to which the link element refers.
21. The computer-readable medium of claim 14, wherein identifying the boilerplate element comprises analyzing a markup language element proximate to at least one of the plurality of elements.
22. A computer-readable medium on which is encoded program code, the program code comprising:
 - program code for identifying a boilerplate element in an article comprising a plurality of elements, the plurality of elements comprising the boilerplate element and a content element;
 - program code for removing the boilerplate element from the article; and
 - program code for indexing the article.
23. A computer-readable medium on which is encoded program code, the program code comprising:

program code for identifying a boilerplate element in an article comprising a plurality of elements, the plurality of elements comprising the boilerplate element and a content element; and

program code for determining a weight for each of the plurality of elements based at least in part on whether the element is a boilerplate element.

24. The computer-readable medium of claim 23, further comprising:

program code for receiving a search query;
program code for determining articles relevant to the search query; and
program code for ranking the articles based as least in part on the determined weights.